Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 33 (2016) 332 – 339

Procedia Environmental Sciences

The 2nd International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring 2015, LISAT-FSEM 2015

Web-based classification application for forest fire data using the shiny framework and the C5.0 algorithm

Gita Puspita Siknun*, Imas Sukaesih Sitanggang

Department of Computer Science, Bogor Agricultural University, Darmaga, Bogor 16680, Indonesia

Abstract

Forest fires are threats for our ecosystems and environment because their impact is very harmful. Every year, the number of hotspots increases, indicating the increase of forest fires in some regions in Indonesia, one of them is Riau Province. To predict the hotspot occurrence, we build a web-based application based on characteristics of area using the Shiny framework. We use the C5.0 algorithm by generating tree and rule-based classification models. The Shiny framework was implemented using reactivity expression, when an input changes, the server will rebuild the output based on the input data. We use the dataset of forest fires in Rokan Hilir district, Riau Province, in 2008. The dataset consists of ten explanatory layers (physical, weather, and socio-economic characteristics) and one target layer (hotspot or non-hotspot). The implementation of the C5.0 algorithm on forest fire data resulted tree models with accuracy of 72.72% and rule-based models with accuracy of 73.51%. The output of tree models is 16 classification rules while the output of rule-based models is 15 classification rules.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015

Keywords: C5.0; classification; forest fire; framework shiny; hotspot

1. Introduction

Forest fires are threats for our ecosystems and environment because their impact is very harmful. According to study in Bappenas and ADB on forest fires, total fire affected land area during El Nino (ENSO) event in 1997/1998

* Corresponding author. Tel.: +6285-6226-6650.
E-mail address: g.puspita27@gmail.com

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015
doi:10.1016/j.proenv.2016.03.084

was about 9.75 million ha [1,2]. The fires emitted much carbon emissions and caused environmental impacts on regional to global scale.

The forest fire prevention is done by monitoring active hotspot data from satellite observations using geographic information system (GIS) to obtain the hotspot occurrences on some areas. Many hotspot occurrences found in logged or drained wetland and peat land areas in some districts in Sumatra, such as Rokan Hilir District, Riau Province. In 2014, there were 85 hotspots in Rokan Hilir District, scattered in several areas [3]. To reduce the incidence of forest fires, it is necessary to build an information system which is applying the data mining techniques. Data mining is able to process and analyze a large amount of data such as forest fire data.

Classification is one of data mining techniques. Previous research about developing classification models on spatial data has been conducted [4] by applying a spatial decision tree algorithm. In the study, the result of spatial decision tree was compared to ID3 and C4.5 algorithms resulted the classification models on non-spatial data. ID3 and C4.5 algorithms that were applied by some researchers have been available in desktop-based application, Weka data mining software. The spatial decision tree algorithm has resulted a higher accuracy of 71.66% than non-spatial classification models using the ID3 and C4.5 algorithm [4]. The ID3 algorithm had accuracy of 49.02% while the accuracy of C4.5 algorithm was 65.24%. Another research was conducted [5] to generate the classification rules using spatial entropy-based decision tree algorithm. The study used the 5-fold cross validation test to split both training set and testing set. However, the classification model on both researches was not implemented into an application that can be used by the related parties.

A recent study has been conducted to develop web-based applications. Hayardisi et al. [6] built a web-based OLAP (On-line Analytical Processing) integrated with a data warehouse for hotspot distribution data. The OLAP application provided the users summarization and visualization in crosstabs and graphs (bar plots and pie plots). Another study from Thariqa and Sitanggang [7] that developed web-based SOLAP (Spatial On-line Analytical Processing) by displaying the summary of hotspots based on socio-economic information. Based on the study, most of hotspots were occurred in a low population density and low school density. However, both studies have not developed an updating menu to handle new hotspot data and the users need to predict the hotspot occurrences based on another characteristics.

A web-based geographic information system (GIS) was developed using OpenGeo Suite for classifying hotspot occurrences [8]. The research applied the C4.5 algorithm in a Java implementation called J48 using Weka data mining software. The application has a menu to predict the hotspot occurrences based on area characteristics with 69.56% accuracy. However, the area characteristics only consisted of the location of the distance to nearest main cities, roads, and rivers. Moreover, if other supporting factors are added in the GIS, the dataset should be re-classify using Weka and the classification rules should be re-implemented in the GIS.

A web-based application using the Shiny Framework was developed using the DBSCAN algorithm [9]. The study performed the clustering method on two dataset of hotspots on Kalimantan Island and South Sumatera Province in 2002-2003. The application displayed the spread pattern of hotspots from the chosen data by user. However it has not implemented a menu to upload a new dataset.

In 1980, JR Quinlan introduced the C4.5 algorithm as the development of ID3 algorithm. Then JR Quinlan was continually developing the classification tree and rule-based models into C5.0 algorithm [10]. C5.0 was proprietary and commercially available until 2011 when a GPL version was released [10].

Previous researches have not implemented the classification models into a web-based application that can be accessed by the related parties to prevent forest fires. Hence, this study developed a web-based application to display classification models and to predict hotspots occurrence. The Shiny framework can build a web-based application which is published on the Internet. It is easily developed and integrated with a web content using HTML and CSS [11]. The Shiny Framework is an R package which is responsive to display the results of the data mining analysis into web-based applications. This web-based application was developed using the R programming language which is open source with several popular packages [12]. In the R package, there are many classification algorithms that can be easily used by users. In this study, the classification algorithm applied is the C5.0 algorithm.